DOCUMENT RESUME

ED 234 039 SP 022 989

AUTHOR Mangano, Nancy G.; And Others
TITLE External Validity Issues Associated with Classroom Observational Research.
INSTITUTION Texas A and M Univ., College Station. Instructional Research Lab.
PUB DATE 25 Jan 83
NOTE 16p.
PUB TYPE Information Analyses (070) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; *Classroom Research; Educational Research; Elementary Secondary Education; Evaluation Methods; *Interaction Process Analysis; *Measurement Techniques; Student Behavior; Teacher Behavior; Teacher Effectiveness; Teaching Methods; *Test Validity

ABSTRACT

This paper examines aspects related to the validity of process-product research involving classroom observation and teacher effectiveness. Differences in categories, terminology, and definitions of five classroom observation instruments are delineated and discussed in relation to their various effects on the validity of process-product research findings. The instruments analyzed are: (1) Teacher and Child Dyadic Interaction observation system; (2) Reading and Mathematics Observation System (RAMOS); (3) Coding System for the First Grade Reading Group Study; (4) Classroom Observation Instrument; and (5) Group Reading Interaction Pattern Observation Instrument (GRIP). Inconsistencies between actual verbal behaviors and teachers' intentions for these verbal behaviors are also discussed. Suggestions for selecting and writing terms and definitions point out the necessity for consistency of terminology and definitions of categories and subcategories across observation instruments. It is recommended that coding of teachers' intended behaviors, rather than their exact linguistic form, would result in more valid findings in the area of process-product research. (Author/JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED234039

External Validity Issues Associated with Classroom Observational Research

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
William H. Rupley
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Nancy G. Mangano[1]
Kansas State University

William H. Rupley
Texas A&M University

Victor L. Willson
Texas A&M University

Running Head: Validity Issues

Submission Date: January 25, 1983

[1]All correspondence will be directed to Nancy G. Mangano, Department of Curriculum & Instruction, Kansas State University, Manhattan, KS 66506

Instructional Research Laboratory
Educational Curriculum & Instruction
Texas A&M University, College Station, TX.
Technical Series # R83005

SP022 189

Abstract

Differences in categories, terminology and definitions of five observation instruments were delineated and discussed in relation to their varying degrees of effect on the validity of the process-product research findings. In addition inconsistencies between actual verbal behaviors and the teacher's intentions for these verbal behaviors were discussed. Finally, suggestions for selecting and writing terms and definitions were delineated and a stand on whether observers should code linguistic behavior or the teacher's purpose for the linguistic behavior was made.

External Validity Issues Associated with Classroom Observational Research

Although a variety of research methods have been used in process-product research, observation instruments are one of the most practical and ecologically valid tools (Snow, 1974) for specifying teacher-student interactions in natural settings. While this is the case, classroom investigations are only as valid as the instruments that are used to measure teacher and pupil behaviors. An instrument is considered to be valid to the extent that it does what it's designed to do. Since the purpose of process-product research is to determine teacher behaviors that enhance student achievement (Heilman, Blair, and Rupley, 1981), instruments used in the classroom must be able to adequately describe interactions between teachers and students that contribute to achievement. Further, the validity of process-product research must be considered in relation to the similarities and differences across instruments used to measure teacher and pupil behaviors.

In an attempt to examine aspects related to the validity of the process-product research, five observation instruments used in teacher effectiveness investigations were compared. The major purpose of this paper was to address how the differences across instruments affect the external validity of this line of research. In addition, inconsistencies between actual linguistic behaviors and the teacher's intended purpose for these behaviors were discussed. Finally, implications for these inconsistencies were highlighted and considerations delineated.

Instrument Selection

The instruments selected for analysis were (1) Teacher and Child Dyadic Interaction observation system (Brophy and Good, 1969), (2) Reading and Mathematics Observation System (Ramos) (Calfee and Calfee, 1975), (3) Coding System for the First Grade Reading Group Study (Brophy, Mahaffey, Greenhalgh, Ogden, and Seilig, 1975), (4) the Classroom Observation Instrument (Stallings, 1980), and (5) Group Reading Interaction Pattern Observation Instrument (GRIP) (Mangano and Rupley, 1982).

Although various systems were initially examined, the above instruments were selected for analyses for several reasons. First, they can be used to observe interactions between teachers and pupils in the classroom. Second, they have been used in process-product research and finally, the authors have previously addressed the reliability and internal validity of the instrument.

The Teacher and Child Dyadic Interaction observation instrument is designed to study interactions between an individual student and the teacher and can be used during any class activity in any content area. The Classroom Observation Instrument can also be used in any subject and its purpose is to provide records of educational processes including teacher behaviors, interactions between teachers and students, and grouping procedures. The Coding System for the First Grade Reading Group Study, as its name implies, was developed for the First Grade Reading Group Study but the authors maintain that the instrument is appropriate for use in primary grade reading classrooms. Its purpose is to measure interactions between individual students and the teacher in

the reading group. The purpose of GRIP is to specify instructional process behaviors used by the teacher during reading instruction. Its focus of observation is the teacher and the pupil with whom the teacher is interacting. The final system under analysis, RAMOS, has various forms. The form that will be discussed in this paper was modified to meet the needs of the Beginning Teacher Evaluation Study. It is intended to measure total time spent in activities related to reading or math, to describe characteristics of reading and math activities, and to delineate the relative distribution of time spent in these activities.

After careful examination of the five instruments, similarities and differences across the instruments were noted and categorized. It was assumed that similarities reflected common hypotheses related to specific process variables across studies. Therefore, these similarities are not discussed at this time. However, differences across instruments have varying degrees of effect on instructional research and may be classified under one of two categories: differences that enhance the validity of the process-product research and differences that limit the external validity or generalizability of this line of research.

Differences that Enhance the Validity of Process-Product Research

Variations in observation instruments that enhance the validity of process-product research are those that result from the conceptualization and formation of research products. These differences allow for specialized instruments that can capture specific aspects of teacher-pupil interactions in the classroom depending on the research problems. Variations are generally reflected in the categories,

subcategories and coding method of the instruments. Evidence of these variations were shown in the categories and subcategories of the aforementioned observation instruments. They can be easily illustrated by the category of "questions or questioning" which was found in all systems analyzed.

The Teacher and Child Dyadic Interaction instrument requires observers to discriminate between questioning subcategories based on whether pupils are asked to exhibit problem-solving behaviors, elicit a single correct response, choose the correct answer, or make a non-academic contribution to classroom discussion. The focus of the question category in the Coding System for the First Grade Reading Group Study requires the observer to discriminate between skill comprehension and comprehension-related questions. The observer records when students are asked to repeat a word just read to them, give an answer to a skill or comprehension question that has a set of four or fewer alternatives, attach a label to a written symbol or answer a question about the sound and letters of words, or break a word or letter down into its component parts. Observer discrimination in this system involves recording information about whether the teacher asks the pupil to relay a personal experience or opinion related to the academic topic.

GRIP also contains multiple subcategories related to questioning. These include the discrimination between teacher-generated questions that call for an academically-related response with only one correct answer and more than one correct answer, nonacademic questions, and rhetorical questions. Further questioning categories include teacher behaviors of probing and restating a question. These latter two

categories are represented under the feedback category in the Teacher and Child Dyadic Interaction instrument and the Coding System for the First Grade Reading Group Study.

Minimal category discrimination is included under question and questioning behaviors in RAMOS and the Classroom Observation Instrument. RAMOS includes one category related to questioning, that defined as the student or teacher giving a direct question and expecting a direct response. The Classroom Observation Instrument includes two types of questions: direct and open-ended. Probing questions are categorized as a special case of feedback.

Differences that Limit the Validity of Process-Product Research

While previous discussion indicated that variations in categories and subcategories of observation systems resulted from the conceptualization and formation of research problems, other differences, in fact, limit the external validity of the research findings. One area where this is evident is in the use of various labels for similar or the same behaviors. For example, teacher behaviors intended to elicit a correct response from a pupil who did not respond correctly is a specific case of sustaining feedback (gives clues) in both the Teacher and Child Dyadic Interaction instrument and the Coding System for the First Grade Reading Group Study, and a special case of corrective feedback (guides) in the Classroom Observation Instrument. It is listed under the questioning category as probes/cues in GRIP, and is not incorporated in RAMOS. A question expected to elicit a single correct response is called a "product" question in the Teacher and Child Dyadic Interaction

system, a product or comprehension question depending on the content of the question in the Coding System for the First Grade Reading Group Study, a direct question in both the Classroom Observation Instrument and GRIP and referred to under the question-answer category in RAMOS.

The use of the same word to describe different behaviors can also limit the generalizability of process-product research findings. An example of this incident is evidenced in the use of the term "direct question." It is defined in the Teacher and Child Dyadic Interaction instrument as an instance in which the teacher calls on a pupil who is not seeking an opportunity to respond. The Classroom Observation Instrument defines this term as a request for direct recall of previously leanred material, while GRIP specifies it as an instance where a teacher or pupil asks a question that has only one correct response. The Coding System for the First Grade Study makes no use of the term "direct instruction," and RAMOS only refers to it under the question and answer category without attaching a definition, i.e., students are given a direct question and are expected to give a direct answer.

These differences have serious implications for the external validity of research findings. It is difficult to discuss and generalize research findings when terms vary across studies. Use of atypical definitions or terminology not only adds to this problem but causes difficulties in replicating the studies. Results may also be unreliable when observers are either unfamiliar with the terms or lack sufficient experience with classroom methodologies to competently use the observation system in question. For example, suppose that an observer who has never taught is coding a direct question on the Classroom

Observation Instrument. The coder may not be aware of verbal signals that allow more experienced observers or one-time teachers to realize that the question is referring to previously learned materials. These incidences may decrease the reliability of the instrument and create deceptive research findings.

Observation and Reality -- Hitting the Bullseye

Inconsistencies between actual teacher behaviors and the teacher's intended purpose for these behaviors can also create recording/coding difficulties for the observer that may jeopardize the validity of process-product studies regardless of the instrument being used. For example, suppose that an observation instrument contains categories for teacher command and teacher questioning behaviors. Consider the following verbal examples used by a classroom teacher who is being observed: "What is the name of the boy in the story?"; "Tell me the name of the boy in the story."; "The name of the boy in the story is _____." All three examples represent an intent to question but only one is an actual question. If the observer were to record the linguistic form of the verbal interactions, he/she would place the first example under the question category, the second verbal behavior under the command category and be unable to place the third item under a category.

The following question arises: Is it more valid to place a verbal item under the exact category that it represents linguistically or under the category that reflects the intention of the verbal behavior? Although the guidelines for these incidences should be provided during training programs so that all observers in one study can be consistent,

they are rarely reported for other groups of researchers to follow. The lack of consistent guidelines for coders to model from one study to the next make studies difficult to replicate and may create contradictory findings when in fact, the results may actually be the same. In the case where a researcher has failed to provide guidelines for coding verbal behaviors that are different than the purpose for their behavior observers may code these behaviors unreliably.

Suggestions for Instrument Developers and Users

The preceding discussion has provided evidence for inconsistencies in terminology and definitions across instruments that make it difficult to generalize findings across process-product studies. Further, the author has addressed differences between actual linguistic terminology and the teacher's intention for using the verbal behavior that can create unreliable results and contradictory findings in the research. More consistency in definitions and terminology is needed across all studies and a decision as to whether linguistic interactions should be coded verbatim or in terms of their intended usage is warranted. While these points reflect the state of the art at present, the following suggestions may be helpful to instrument developers and users in the future.

Selection of terms and definitions may become more consistent if observation instrument developers review previously developed systems in the area of their research and choose those terms and definitions that are most widely used across studies. In the case where various definitions are noted, those that are the simplest and most logically

understood by the observers should be adopted because they would increase the reliability of the instrument and affect the validity of the findings. If definitions for terms are not found, consulting the literature or educational dictionaries can add to the consistency of the meaning of terms. Well-conceptualized operational definitions for each category of behavior should also be specified. While the latter point may appear obvious and commonplace, operational definitions have been vague in the past (Herbert and Attridge, 1975).

Finally, while instrument categories and subcategories should be as low in the degree of inference as the situation warrants (Herbert and Attridge, 1975), high inference variables may be desirable for giving greater insights into those process behaviors that result in subjectively high ratings of teacher performance (Rosenshine and Furst, 1973; Dunkin and Biddle, 1974). A special problem may arise when defining these variables however. When defining high-inference variables, operational definitions that specify low-inference behaviors to describe the variable should be incorporated (Herbert and Attridge, 1975). For example, in the case of "student involvement" definitions could include the overt behaviors of focusing eyes on materials, writing when students are supposed to be writing, and the like.

A final point in this paper concerns whether it is more valid to code the linguistic behavior or the teacher's intention for the verbal behavior during classroom observations. Examples of verbal behaviors that can be confusing include: "Will you turn to page 36?"; "Clyde?"; "We will walk quietly to our desks, won't we?"; and "Tell me what the next word is." If researchers wish to capture sequences of teacher

behaviors, the coding of the intended behavior rather than the linguistic behavior would seem more valid for the following reasons. First, intended behaviors more clearly reflect what the teacher is trying to accomplish in the classroom. For example, consider the following scenario:

Teacher: Tell me the name of the boy in the story.
Student: John
Teacher: John was the father's name. The boy was named after the grandfather. Now, tell me the name of the boy in the story.
Student: Joseph
Teacher: That's correct.

If observers coded the linguistic form, the interaction would be recorded as the teacher makes a verbal command, student responds incorrectly, teacher gives information and makes another verbal command, student responds correctly, and the teacher gives positive academic feedback. However, coding intended behaviors more clearly provides insight into the situation. The interactions would be read as: the teacher asks a question, student responds incorrectly, teacher gives clue, teacher restates question, student responds correctly, teacher gives positive academic feedback.

Finally, if the major purpose for performing classroom research is to determine strategies and sequences of behaviors that enhance student achievement, it would follow that intended behaviors would be more valid as it eliminates variations in research results that reflect the teacher's style of interacting. Results derived from interactions of intended behaviors would also transfer more directly into the training of preservice and inservice teachers who have their own style of teaching. On the other hand, if the study is linguistic in

nature and the research problem reflects teacher styles in disseminating types of information such as questions, praise, or disciplinary feedback, researchers would profit more thoroughly from studying the actual linguistic form in relation to the intended verbal behavior. In this case, instruments that reflect both types of verbal behaviors would need to be developed.

Summary

More careful attention needs to be given to the external validity of observation instruments used in process-product research in order for the findings from teacher effectiveness research to be more truly generalizable to both preservice and inservice training programs. In summary, terminology and definitions of categories and subcategories must become more consistent across observation instruments. This can begin to result if researchers follow the suggestions delineated above. In addition, if the purpose for teacher effectiveness research is to specify teacher processes that enhance pupil learning regardless of the teacher's personal style, the coding of the teacher's intended behaviors rather than their exact linguistic form would appear more valid.

Teacher effectiveness research has come a long way since its onset. Consistency in measuring classroom interactions will further refine this line of research and result in more valid findings in the future.

References

Brophy, J.E. & Good, T.L. Teacher-Child Dyadic Interaction: A Manual for Coding Classroom Behavior, (Report Series No. 27), Research and Development Center for Teacher Education, The University of Texas at Austin, 1969.

Brophy, J.E., Mahaffey, L., Greenhalgh, C. Ogden, J. Selig, H.M. Coding System for the First Grade Reading Group Study. (Report No. 75-2), Research and Development Center for Teacher Education, The University of Texas at Austin, 1975.

Calfee, R., & Calfee, K.H. Beginning Teacher Evaluation Study: Phase II (1973-74) Final Report: Volume III-2: Reading and Mathematics Observation System: Description and Analysis of Time Expenditures. (ED 127 367), 1976.

Dunkin, M.J. & Biddle, B.M. The Study of Teaching. New York: Holt Rinehart and Winston, Inc., 1974.

Heilman, A., Blair, T., & Rupley, W. Principles and Practices of Teaching Reading: Fifth Edition. Columbus: Charles E. Merrill Publishing Co., 1980.

Herbert, J., & Attridge, C. A guide for developers and users of observation systems and manuals. American Educational Research Journal, 1969, 6, 227-39.

Mangano, N.G. & Rupley, W.H. Group Reading Interaction Pattern Observation Instrument. (Report No. 84003) College Station: Texas A&M University, Instructional Research Laboratory, 1982.

Rosenshine, B.V. & Furst, N. The use of direct observation to study teaching. In P.M.W. Travers (eds.) Second handbook of research on teaching. Chicago: Rand McNally and Co., 1973, 122-182.

Snow, R. Representative and quasi-representative designs for research on teaching. Review of Educational Research, 1974, 44, 265-291.

Stallings, J.A. Learning to Look: A Handbook on Classroom Observation and Teaching Models. Belmont: Wadsworth Publishing Co., 1977.